Univariate Single- and Mutiple-Use Bayesian Linear Calibration

Mark Vangel
Dana-Farber Cancer Institute
Department of Biostatistical Science

Bayesian Working Group
Harvard School of Public Health
March 14, 2001

0-0

Outline

- Overview of calibration, with motivating examples.
- Simple linear calibration: frequentist controversy, Hoadley's Bayesian resolution revisited.
- Multiple-use calibration: curious features illustrated by an example.
- Noninformative priors in polynomial calibration problems

Calibration Problems

- In a typical calibration problem, there are two ways of measuring something.
  - For a "training" dataset, we use both measurements on each object.
  - For future data, we only perform one of the methods, and we use this value, a model, and the training data to make inference on the other measurement method.
- Typically one of the measurement methods is more accurate, but also more expensive, time consuming, etc.
- Sometimes one of the measurements can only be made for the training data because of the use of standard objects, for which the "truth" is known by design.
- Calibration can often be thought of as *inverse regression*.

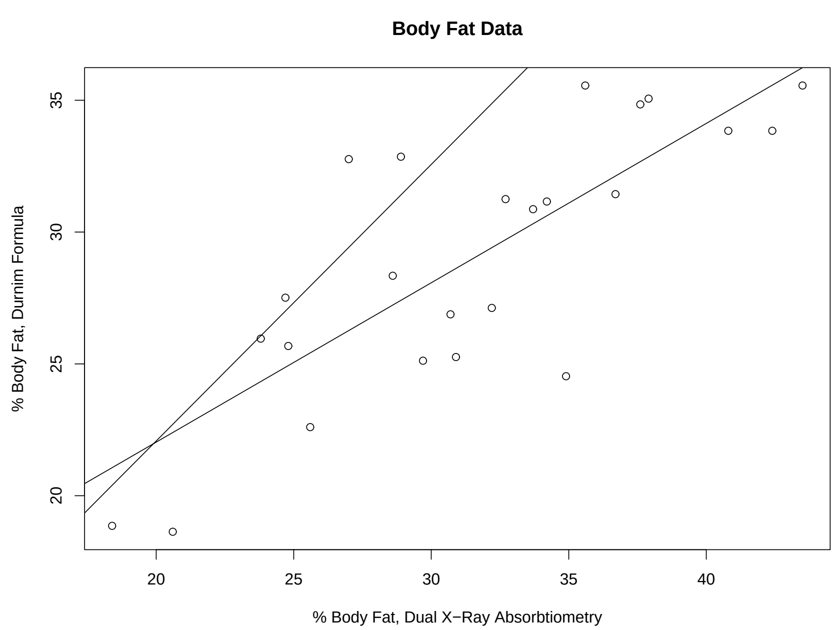
Types of Calibration Problems

- Simple linear ($y = \alpha + \beta x + \epsilon$), linear ($y = X\beta + \epsilon$), nonlinear ($y = f(x_1, \ldots, x_p; \beta) + \epsilon$).
- Univariate or multivariate
- *random* (training xs from population) or *controlled* (training xs known).
- Single-use or multiple-use

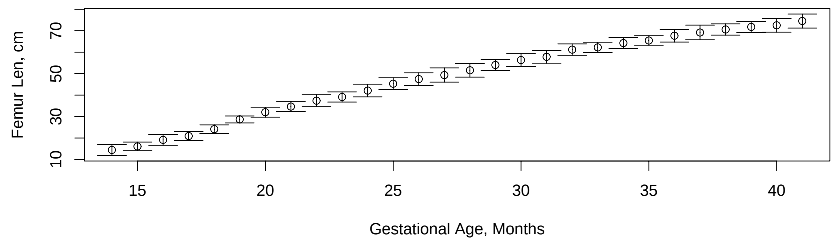
References

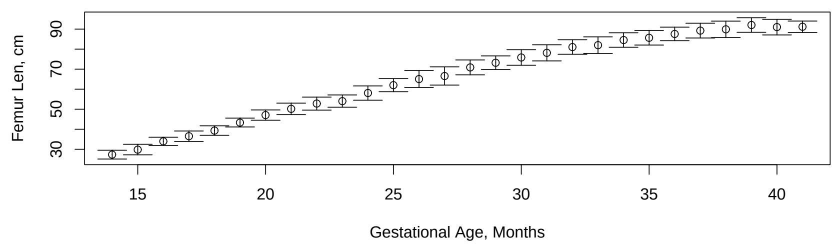
- Eisenhart, C.E. (1939). "The Interpretation of Certain Regression Methods and Their use in Biological and Industrial Research," *Annals Math. Stat.*, **10**, 162-186.
- Krutchkoff, R.G. (1967). "Classical and Inverse Methods of Calibration," *Technometrics*, **9**, 425-439.
- Hoadley, B. (1970). "A Bayesian Look at Inverse Linear Regression," *JASA*, **65**, 356-369.
- Scheffé, H. (1973). "A Statistical Theory of Calibration," *Annals of Stat.*, **1**, 1-37.
- Brown, P.J. (1982). "Multivariate Calibration," (with discussion), *JRSSB*, **44**, 287-321.
- Racine-Poon, A. (1988). "A Bayesian Approach to Nonlinear Calibration Problems," *JASA*, **83**, 650-656.
- Osborne, C. (1991). "Statistical Calibration: A Review," *Int. Stat. Rev.*, **59**, 309-336.

Calibration of Body Fat Measurements (Branco, et al. (2000), JSPI, 90, p. 83)

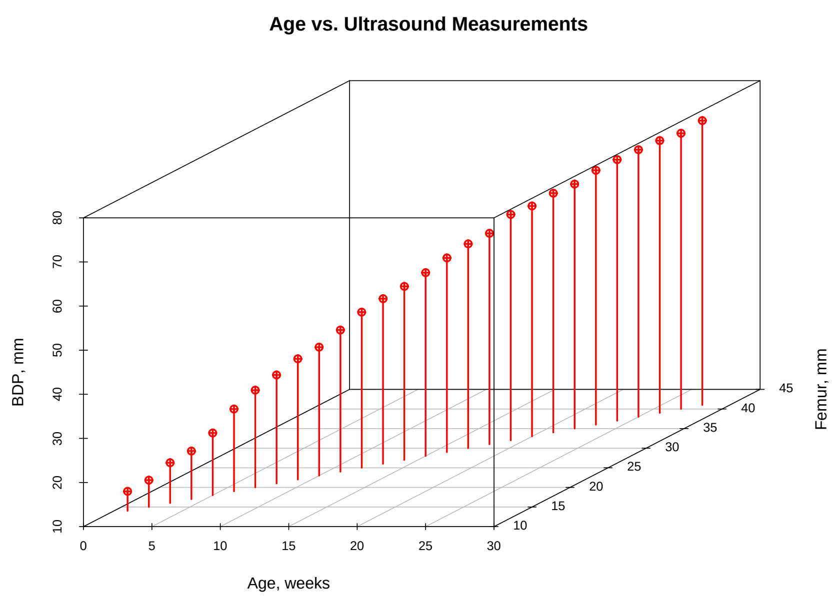


Estimating Gestational Age from Ultrasound Fetal Measurements

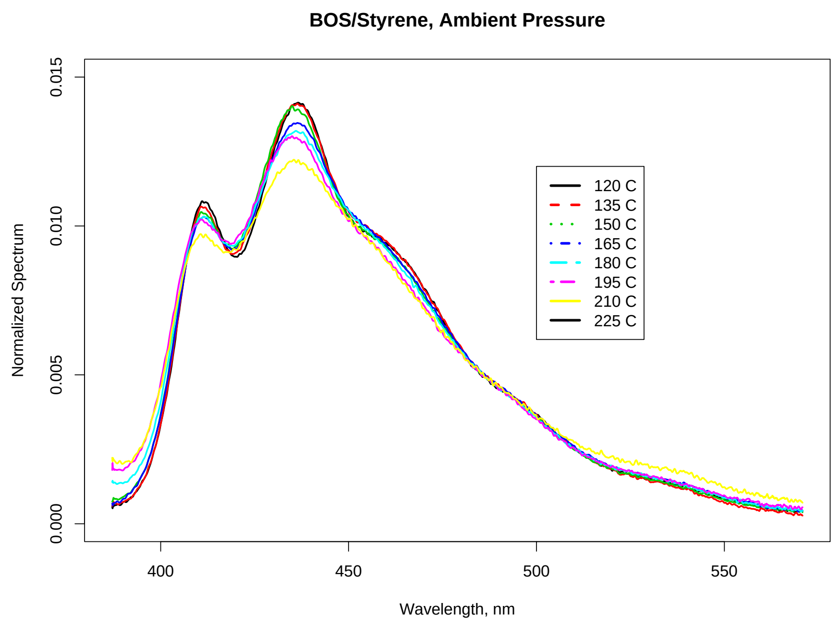




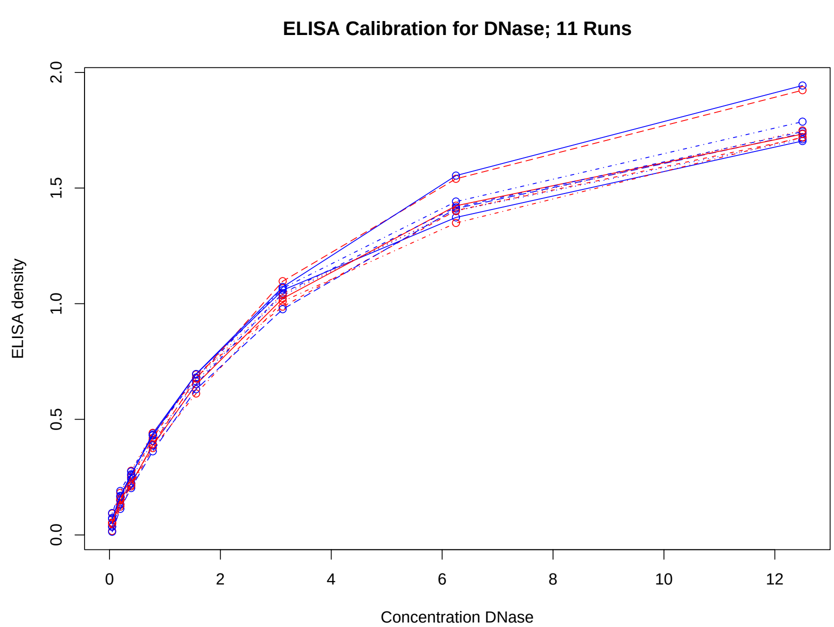
Fetal Measurements vs. Age: Projections of a Curve in R^3



Multivariate Calibration; Wavelength Selection: Estimating Temperature from Polymer Fluorescence Spectra



Developing ELISA Assay for DNase (Davidian and Glitinan, 1995, p. 134)



Univariate Single-Use Calibration:
A Frequentist Controversy With a Bayesian Resolution

- Classical Calibration: Eisenhart (1939)
- Inverse Calibration: Krutchkoff (1967)
- Bayesian Interpretation: Hoadley (1970)

The Simple Linear Regression Calibration Problem

- Training Data:

$$y_i = \alpha + \beta x_i + \epsilon_i,$$

  for $i = 1, \ldots, n$, with $\epsilon \sim \mathrm{N}(0, \sigma^2)$. The $\{y_i\}$ are typically easy and cheap to obtain; the $\{x_i\}$ are more expensive but more accurate.

- Future observations:

$$y_i = \alpha + \beta \eta + \epsilon_i,$$

  $i = n + 1, \ldots, n + m$.

- How should one estimate η, and assess the uncertainty in the estimate?

“Classical” Calibration
Eisenhart (1939)

- Estimate $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$ by least squares, then let

$$\hat{\eta} = \frac{\bar{y}_2 - \hat{\alpha}}{\hat{\beta}},$$

where $\bar{y}_2$ is the mean of $\{y_i\}_{i=n+1}^{m+n}$.

- A confidence interval is obtained by solving a quadratic equation for the values of η which satisfy

$$\frac{|\bar{y}_2 - \hat{\alpha} - \hat{\beta}\eta|}{c(\eta)\hat{\sigma}} \leq t_{n-2}(\gamma),$$

where $c(\eta)$ is the square root of the usual quadratic form for prediction intervals in simple linear regression.

The Classical Calibration Estimator Does not Have Positive Moments

- The estimator $\hat{\eta}$ has no positive moments.
- This follows from the fact that $\hat{\beta}$ has a minimum greater than zero on any interval containing the origin.
- In practice, if $\hat{\beta}$ is many standard deviations from zero, this may not be very important.
- In a sense, one might regard this infinite expectation as arising from a breakdown of the normal model, which one doesn’t really expect to hold far into the tails.

Confidence Regions based on the Classical Estimator are not Always Intervals

- When the "signal-to-noise" ratio β/σ is small, a confidence region can be empty, the union of two rays, or the whole line.
- This is because one obtains this confidence region by solving a quadratic equation, and a parabola can intersect the η axis in various ways.
- The estimator $\hat{\eta}$ is inadmissible (Kubokawa and Robert, 1994).

Inverse Calibration
Krutchkoff (1967)

- Training data:

$$x_i = \beta_0^* + \beta_1^* y_i + \epsilon_i,$$

  for $i = 1, \ldots, n$,
- Future observation:

$$\eta = \beta_0^* + \beta_1^* y_i + \epsilon_i,$$

- Calibration estimate:

$$\check{\eta} = \hat{\beta}_0^* + \hat{\beta}_1^* \bar{y}_2.$$

  Use prediction interval for uncertainty on $\check{\eta}$.

Comments on the "Inverse" Estimator

- "Justified" on basis of MSE comparison with classical estimator, which doesn't make much sense (noted by Williams, 1967).

  > "Although a mathematical proof is not given, the Monte Carlo results are such that one can safely conclude that the Inverse approach to the calibration problem has a uniformly smaller MSE than the Classical approach" (!) [Krutchkoff]

- Inconsistent (Berkson, 1969).
- Bayesian justification with t-prior on η (Hoadley, 1970).
- Admissible (Kubokawa and Robert, 1994).

Controversy

- Krutchkoff (1967) claimed that the inverse estimator had uniformly smaller MSE and so should be preferred.
- Obviously his claim was flawed because $\hat{\eta}$ has infinite variance. (It also has infinite mean, but this was overlooked for some reason in the early literature.)
- But $\check{\eta}$ seemed to work well in simulations. It was easier to use, and didn't have the peculiar confidence interval properties of $\hat{\eta}$.

The Posterior Predictive Distribution in Bayesian Calibration

The following derivation is for *controlled* calibration. (A similar result can be shown for the random calibration case.)

$$
\begin{aligned}
&p(\eta, \beta, \sigma | X_1, Y_1, Y_2) \propto \\
&p(Y_1, Y_2 | X_1, \eta, \beta, \sigma) \cdot p(\eta, \beta, \sigma) = \\
&p(Y_2 | Y_1, X_1, \eta, \beta, \sigma) \cdot p(Y_1 | X_1, \eta, \beta, \sigma) \cdot p(\beta, \sigma) \cdot p(\eta) \propto \\
&p(Y_2 | Y_1, X_1, \eta, \beta, \sigma) \cdot p(\beta, \sigma | Y_1, X_1) \cdot p(\eta)
\end{aligned}
$$

Notation for a Univariate Calibration Model

$$
\begin{aligned}
Y_1 &= X\beta + \epsilon_1 \\
Y_2 &= Z\beta + \epsilon_2 \\
\epsilon_1 &\sim \mathrm{N}\left(0, \sigma^2 I_n\right) \\
\epsilon_2 &\sim \mathrm{N}\left(0, \sigma^2 I_m\right) \\
Z_{m\times p} &= J_n \phi^T \\
\phi &= \left[\begin{array}{cccc} 1 & \phi_1(\eta) & \cdots & \phi_{p-1}(\eta) \end{array}\right]^T
\end{aligned}
$$

- The matrix $X_{n\times p}$ is known, of full rank (for convenience), with a column, J_n, of ones. The functions $\phi_j(\eta)$ are known, and most likely monotone.

Hoadley's Results

- Hoadley (JASA, 1970) was apparently the first to look at the *simple* linear regression calibration problem from a Bayesian perspective.
- He showed that

$$p(\eta|Y_1, Y_2) \propto \frac{p(\eta)}{s\rho(\eta)} t_\nu \left[\frac{\bar{y}_2 - \hat{\beta}_0 - \hat{\beta}_1 \eta}{s\rho(\eta)} \right]$$

- Where

$$\begin{aligned} \rho(\eta) &= \sqrt{\frac{1}{m} + \phi^T (X^T X)^{-1} \phi} \\ \phi &= \begin{bmatrix} 1 & \eta \end{bmatrix} \\ s^2 &= (\mathrm{SS}_R + \mathrm{SS}_{y_2})/\nu \\ \nu &= n + m - 3 \end{aligned}$$

Bayesian Interpretation of the Inverse Estimator

- Hoadley (1970) notes that (for centered and scaled x), if

$$p(\eta) = t_{n-p-1}\left(\eta; 0, \frac{n+1}{n-3}\right)$$

- Then

$$p(\eta|Y_1, y_2) \propto t_{n-2}(\eta; 0, 1)$$

Priors on η

- Hoadley noted that simple linear regression calibration with the usual improper noninformative prior on η leads to an improper posterior.
- This seems to be the Bayesian counterpart to the lack of positive moments of the Classical estimator.
- Also in common with the Classical frequentist approach, the improper posterior is not likely to cause problems unless the training data are very noisy.
- If we integrate numerically, then we'll probably use a finite support for η, anyway.
- If we use simulation, then the improper posterior might arise from humongous values which occur *very* rarely.

Priors on η (Cont'd)

- Note that, if we consider a regression model in η, then Jeffrey's prior on η will lead to a *proper* posterior if the degree of the polynomial is greater than one.
- But if the polynomial is of second or higher degree, than Jeffrey's prior on η will almost always lead to a multimodal posterior. We'll see an example of this later.
- A reference prior for the simple linear regression calibration problem is available, and it will always lead to a proper posterior.

Kubokawa and Robert
(J. Multivariate Analysis, 1994)

$$Y_i = \alpha + \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_d \end{bmatrix} x_i + \epsilon_i,$$

$$\begin{aligned} \epsilon_i &\sim \mathrm{N}_d(0, \sigma^2 I) \\ \hat{\beta} &\sim \mathrm{N}_d\left[\beta, \frac{\sigma^2}{s_{xx}}\right] \end{aligned}$$

- Note that α and x_i are scalars, $i = 1, \ldots, n$.
- When $d = 1$, we have simple liner regression. I'm not sure if $d > 1$ is of much use in applications.

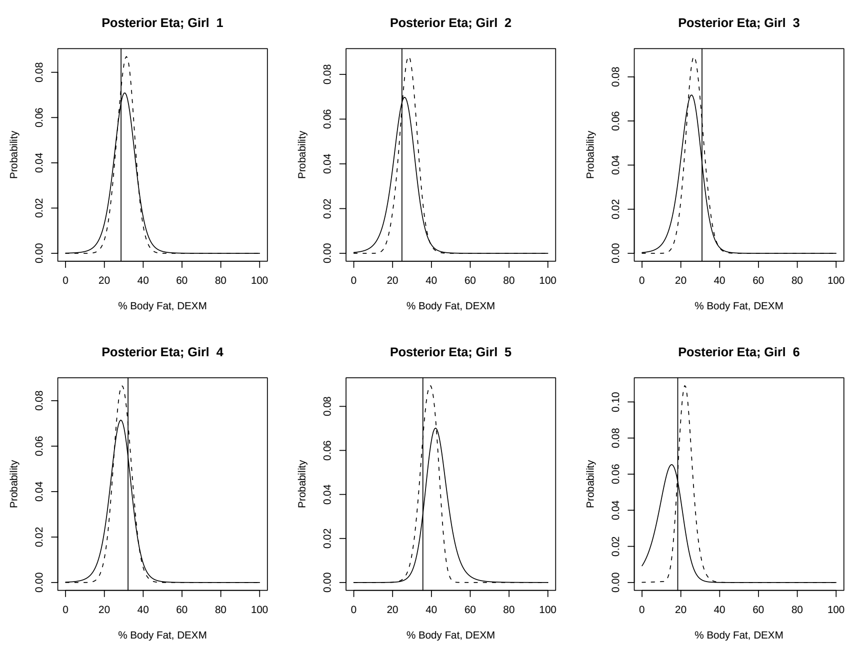
Kubokawa and Robert:
Reference Prior

- For calibration with this model (in particular, for simple linear regression calibration), the Jeffrey's prior on η leads to an improper posterior.
- The authors derive the reference prior for this problem, for which the posterior is proper. When $d = 1$, this prior is, in our notation,

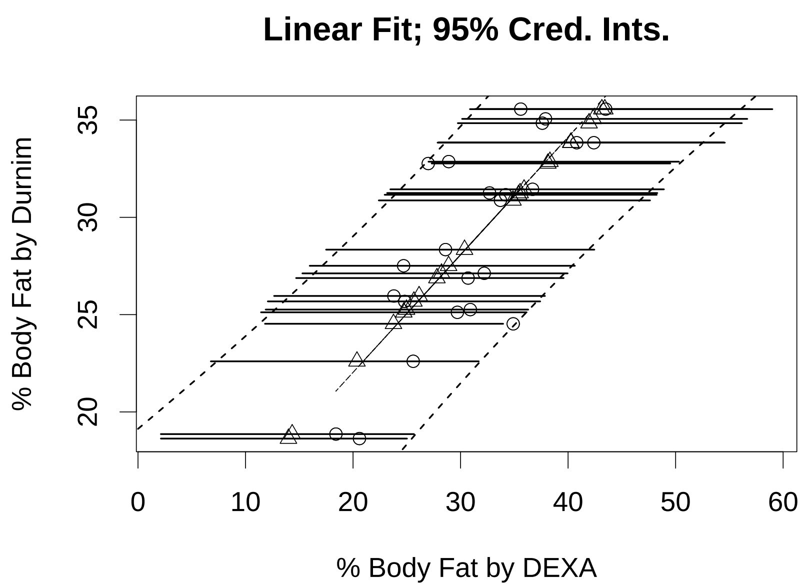
$$p(\eta, \beta, \sigma^2) \propto \frac{\eta}{(\sigma^2)^{3/2}}$$

First Numerical Experiment: Bodyfat Example; Single Use

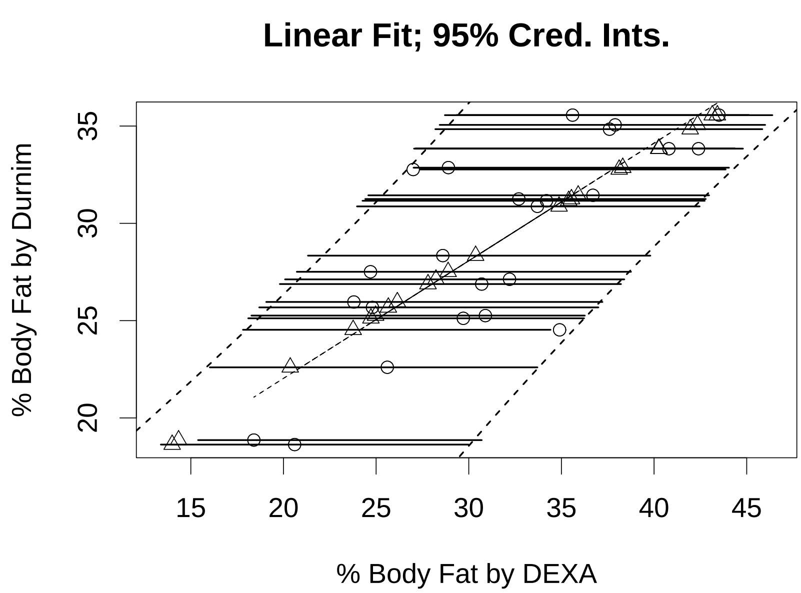
- Leave each y_i out of the training set in turn, determine the posterior of the corresponding η_i, and compare with the omitted x_i.
- Calculate 95% credible intervals in this way for each x_i, and compare with Classical and Inverse estimates.
- **Result:** For **controlled** calibration the intervals agree with the *Classical* approach; for **random** calibration, the intervals agree with the *Inverse* approach.



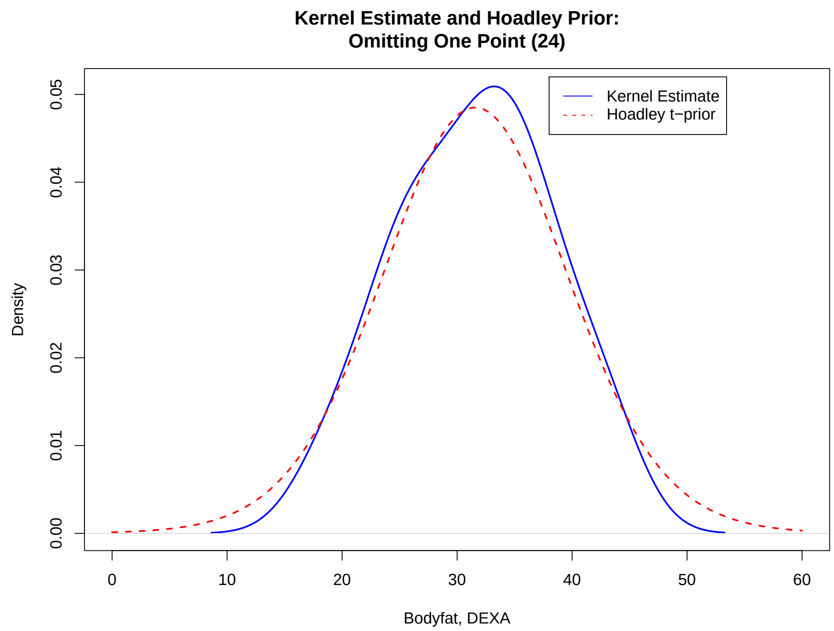
“Leave-One-Out” Example (Controlled Calibration): Broken Lines Indicate Classical Intervals



“Leave-One-Out” Example (Random Calibration): Broken Lines Indicate Inverse Intervals



Kernel Density and Hoadley Prior Nearly Identical



Frequentist Multiple-Use Calibration Sechffé (*Annals*, vol 1., p. 1, 1973)

- Most calibration curves are used more than once; often many times.
- The standard frequentist approach to multiple-use intervals is due to Scheffé (1973).
- This approach introduces a second probability, with confidence statements which are (approximately) of the form “for γ_1% of all training data sets, for γ_2% of all future ys, the corresponding “true” η will be in the calibration interval”
- The precise confidence statement appropriate for Scheffé’s interval is difficult to state concisely; usually the above approximation is used instead.

Bayesian Multiple-Use Calibration

- Assume that, having observed post-training y_j^*s, $y_1^*, \ldots, y_m^*$, we want to determine the joint posterior of the corresponding η_js.
- We can determine the conditional distribution of any η_i, given all the other η_js, from the posterior predictive of y_i^*, regarded as a function of η_i^*, and treating the points (η_j, y_j^*), $j \neq i$, as "data".
- This requires an easy one-dimensional integration.
- From these full conditionals, we can then determine the joint posterior of $\{\eta_j\}_{j=1}^m$ using Gibbs sampling.
- (Alternatively, of course, we can use Gibbs [and Bugs] to do the whole job.)

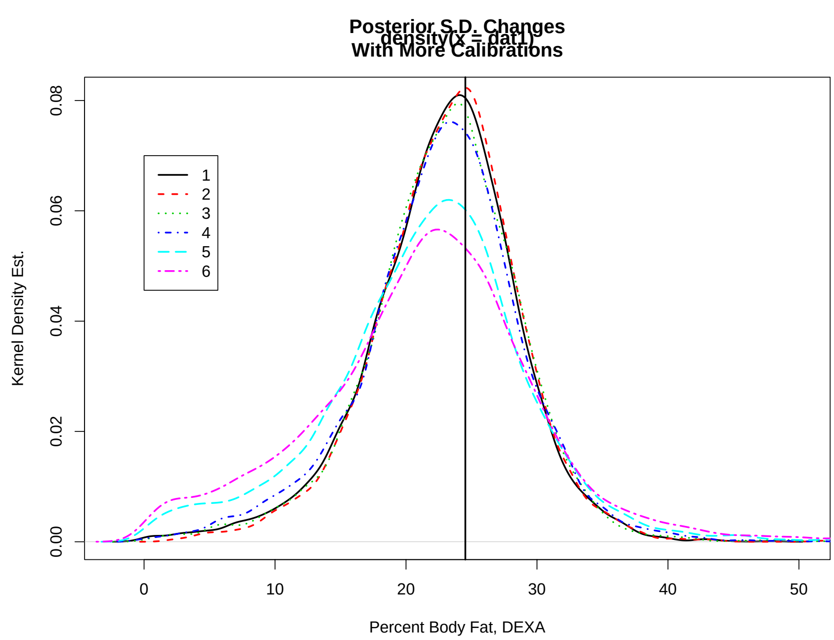
Some Characteristics of This Bayesian Approach

- The calibration results, posteriors of $\eta_1, \ldots, \eta_m$, are *not* dependent; observing a new y_{m+1}^* changes all earlier calibration inferences.
- The posteriors model parameters (β, σ) also change with each new y^*.
- At least for controlled calibration, the posterior of β can change substantially. This is at first counter-intuitive, since we observe no xs after the training data!

Second Numerical Experiment: Bodyfat Data; Multiple-Use

- Use the data on the first 18 girls as training data, and girls 19-24 for calibration.
- Look at the joint posterior of η for cases 19-24, for both controlled and random calibration.
- Take one of the "calibrated" girls (#19), and see how the posterior for η changes depending on what other calibrations have been performed.
- **Result:** The posterior of β can change considerably in repeated use controlled calibration. This is apparently because some of the training data becomes less relevant if it is not in the range of the multiple y_j∗s.

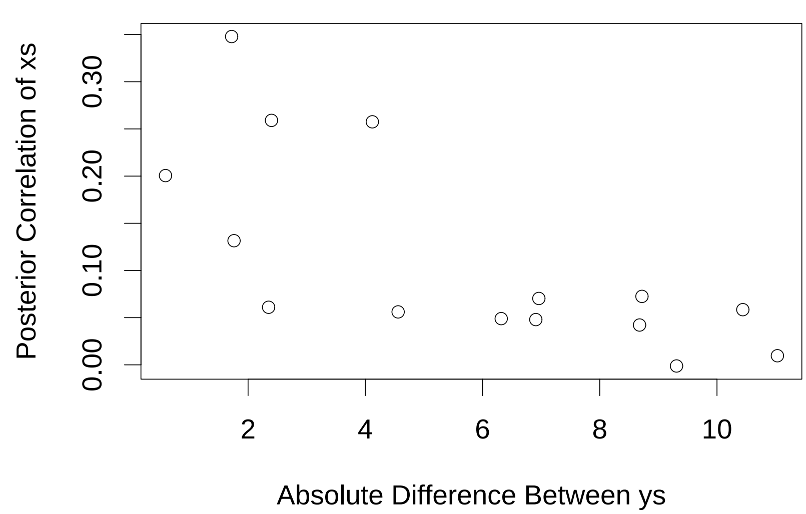
Bodyfat Data, Controlled Calibration Posterior of η_1 Given $\{y_1^*\}, \ldots \{y_1^*, \ldots, y_6^*\}$

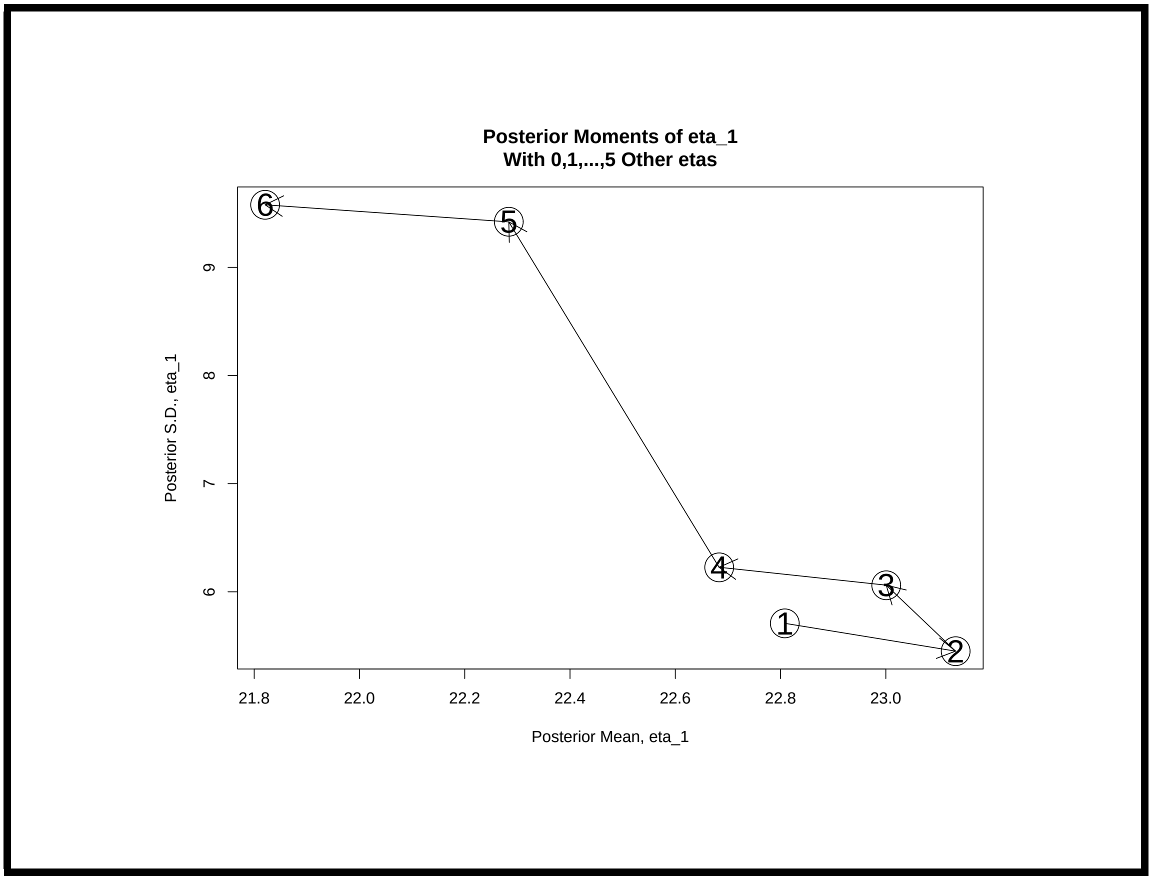


Posterior Correlations of η_is
Bodyfat Data, Controlled Calibration

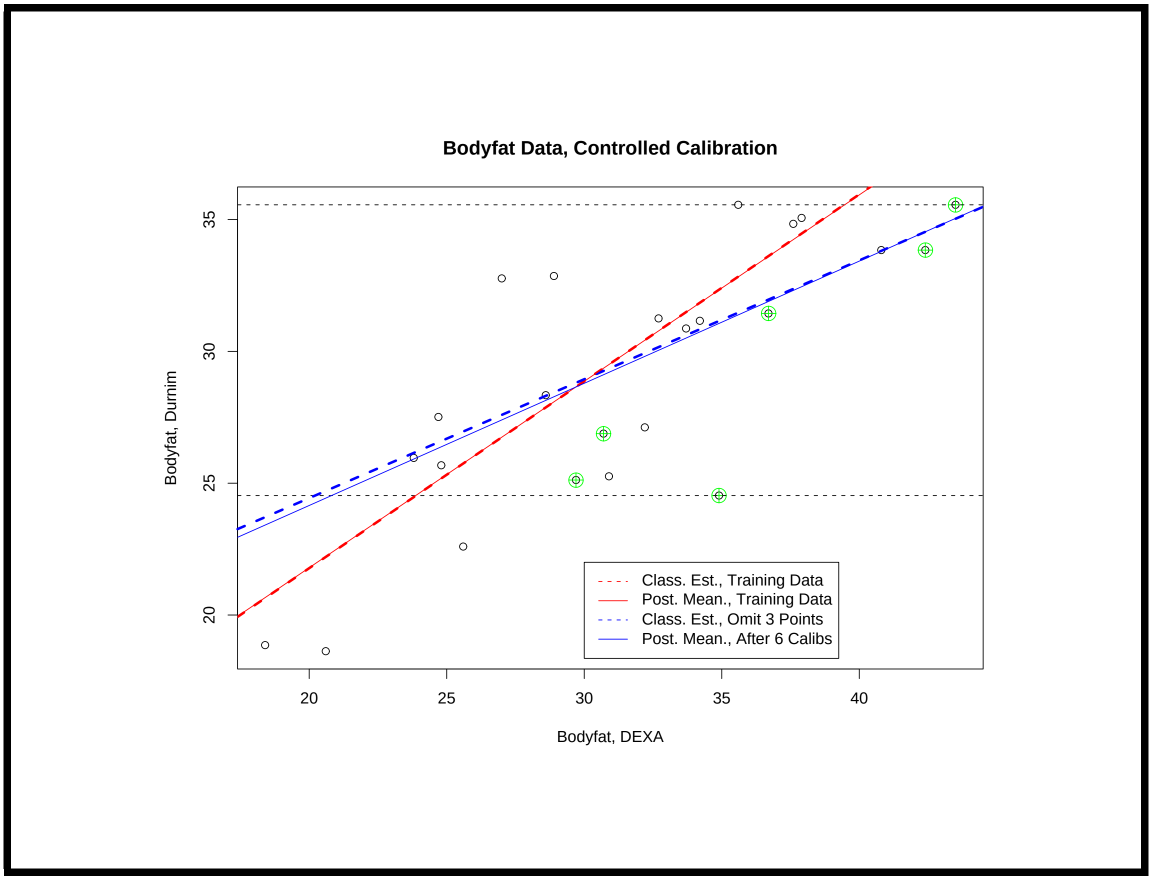
$$\begin{bmatrix} 1 & 0.06 & 0.05 & 0.00 & 0.01 & 0.20 \\ 0.06 & 1 & 0.06 & 0.07 & 0.04 & 0.13 \\ 0.05 & 0.06 & 1 & 0.26 & 0.26 & 0.05 \\ 0.00 & 0.07 & 0.26 & 1 & 0.35 & 0.07 \\ 0.01 & 0.04 & 0.26 & 0.35 & 1 & 0.06 \\ 0.20 & 0.13 & 0.05 & 0.07 & 0.06 & 1 \end{bmatrix}$$

Posterior Correlations of η_is
Bodyfat Data, Controlled Calibration

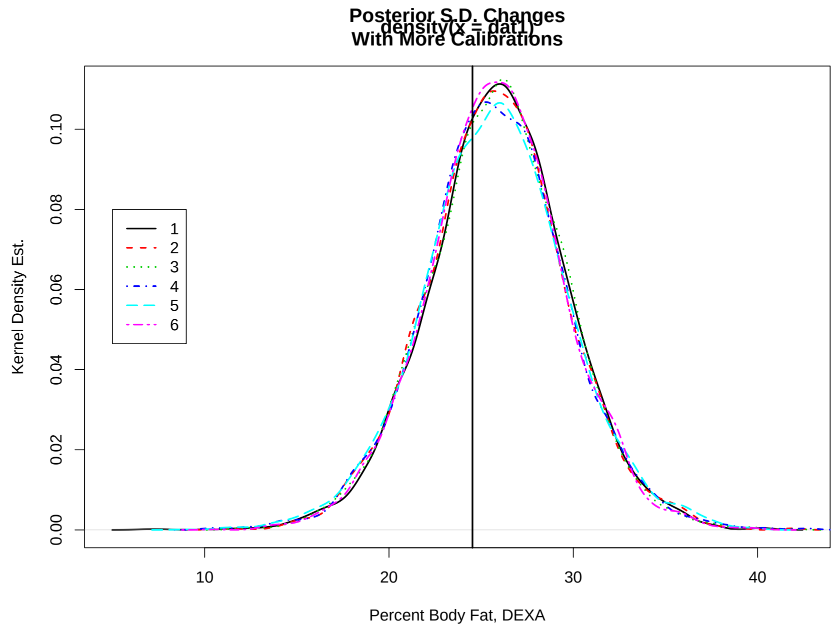



Posterior Moments of eta_1
With 0,1,...,5 Other etas
Posterior S.D., eta_1
Posterior Mean, eta_1
21.8
22.0
22.2
22.4
22.6
22.8
23.0
6
7
8
9
1
2
3
4
5
6

38


Bodyfat Data, Controlled Calibration
Bodyfat, Durnim
Bodyfat, DEXA
20
25
30
35
40
Class. Est., Training Data
Post. Mean., Training Data
Class. Est., Omit 3 Points
Post. Mean., After 6 Calibs

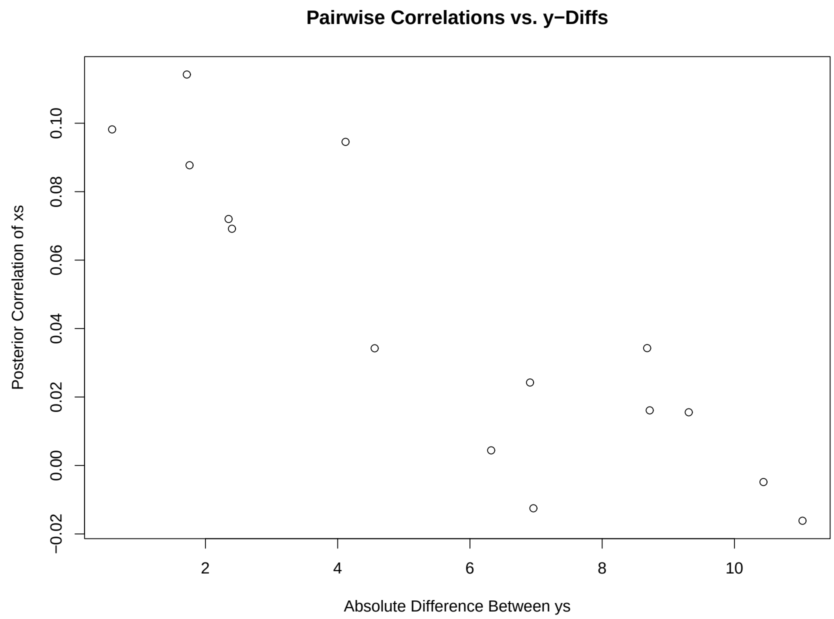
Bodyfat Data, Random Calibration
Posterior of η_1 Given $\{y_1^*\}, \ldots \{y_1^*, \ldots, y_6^*\}$



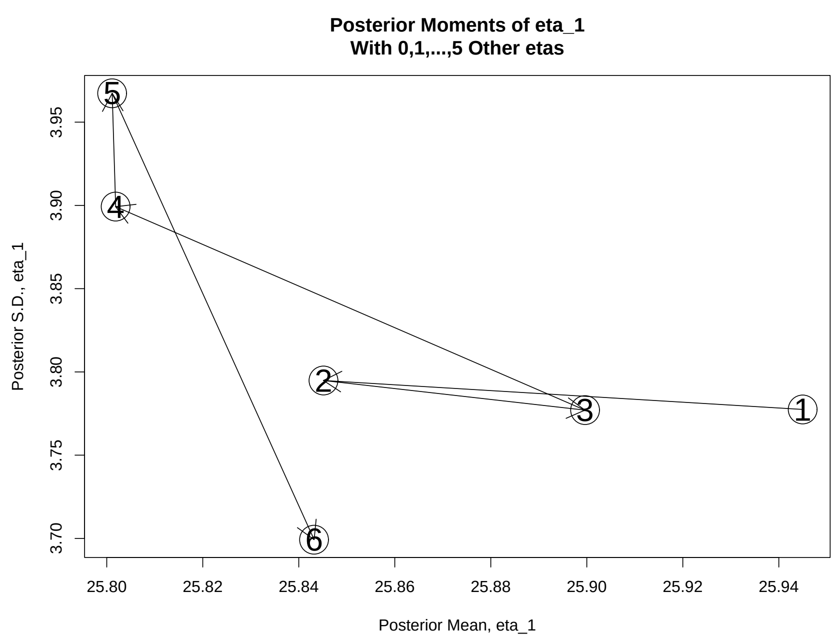
Posterior Correlations of η_is
Bodyfat Data, Random Calibration

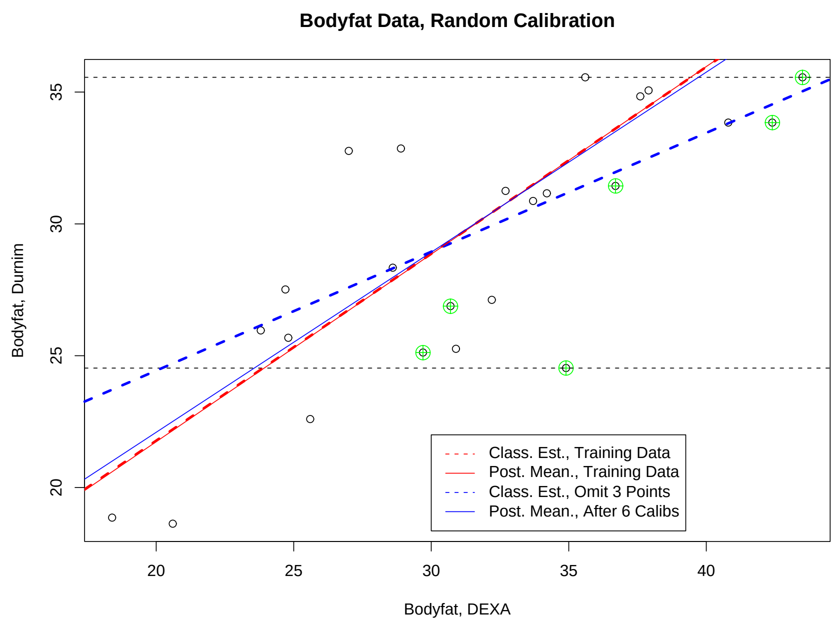
$$\begin{bmatrix} 1 & 0.07 & 0.02 & 0.02 & -0.02 & 0.10 \\ 0.07 & 1 & 0.03 & -0.01 & 0.03 & 0.09 \\ 0.02 & 0.03 & 1 & 0.07 & 0.09 & 0.00 \\ 0.02 & -0.01 & 0.07 & 1 & 0.11 & 0.02 \\ -0.02 & 0.03 & 0.09 & 0.11 & 1 & 0.00 \\ 0.10 & 0.09 & 0.00 & 0.02 & 0.00 & 1 \end{bmatrix}$$

Posterior Correlations of η_is
Bodyfat Data, Random Calibration

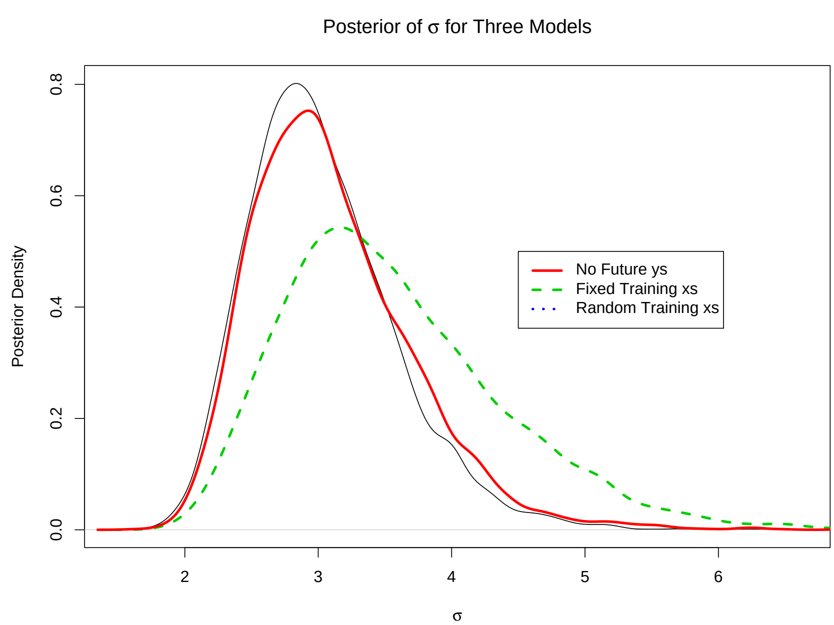


42





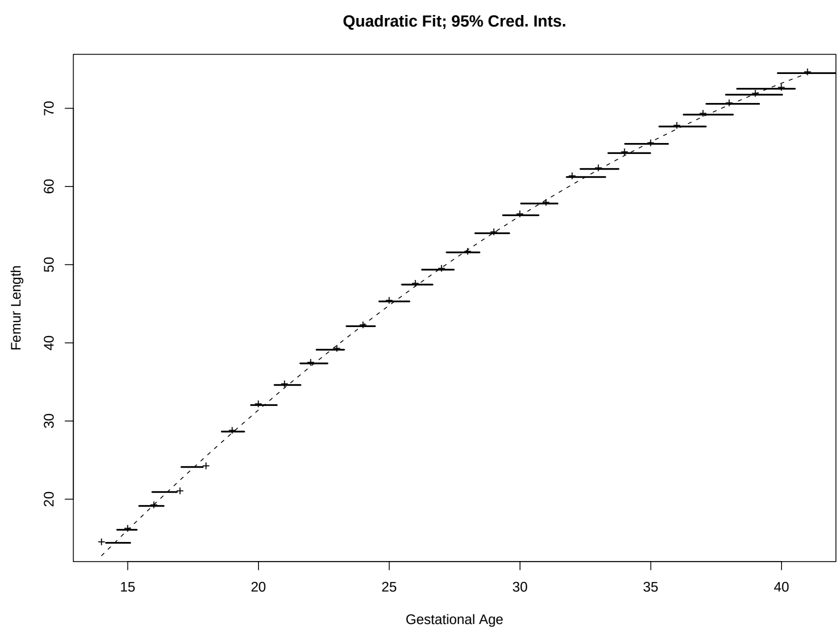
Calibration Data Provides Information on σ



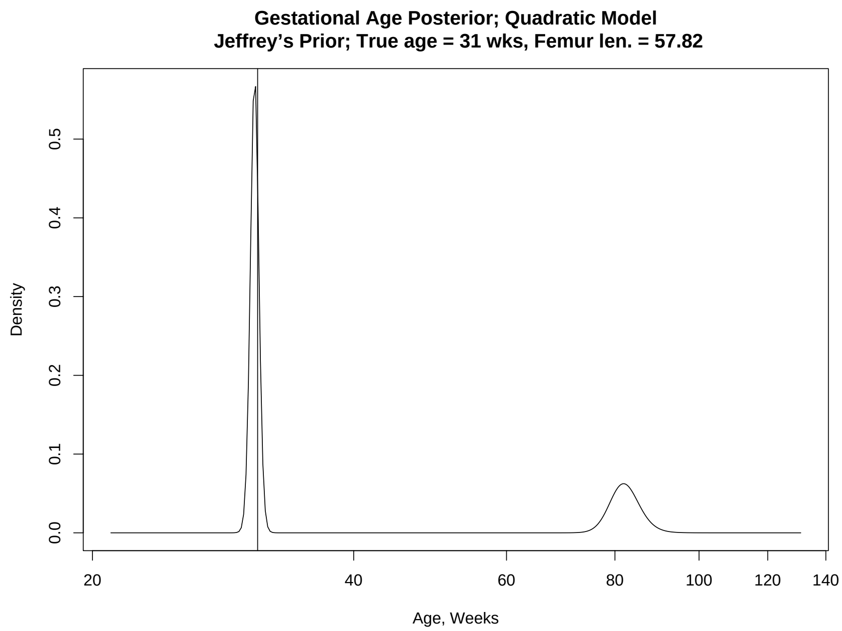
Third Example: Gestational Age

- This example is mostly redundant; I include it here to show how even when the usual noninformative prior on η leads to a proper posterior, it may not be of use because this posterior might not make sense.
- Multimodal posteriors for η will typically arise from polynomial models of degree > 1 because $f^{-1}(y)$ will usually not be unique!

A Quadratic Bayesian Calibration



Use a Proper $p(\eta)$ When $f^{-1}(y)$ can Have Multiple Values



Summary of Main Results

- A Bayesian perspective sheds much light on the Classical/Inverse calibration controversy.
  - The Inverse estimator has a Bayesian interpretation, and is preferable when future ηs will lie within range of training data.
  - The Inverse estimator is also better from a Bayesian point of view in the random calibration case.
- A Bayesian approach shows particular promise in the multiple-use setting. One interesting new feature which arises is that the posterior can "reweight" the training data after several post-training ys have been observed.